LEVEL ~ (4)

Research Memorandum 74-6

EVALUATION OF HIGH DENSITY FORMAT
FOR AFQT ANSWER SHEET

D D C
RECEIVED
NOV 15 1979
E

ari

U. S. Army

Research Institute for the Behavioral and Social Sciences

April 1974

79 11 13 313

DISPOSITION FORM

For use of this form, see AR 340-15, the proponent agency is TAGCEN.

REFERENCE OR OFFICE SYMBOL	SUBJECT
PERI-TP	Clearance and Transmittal of Reports to DTIC

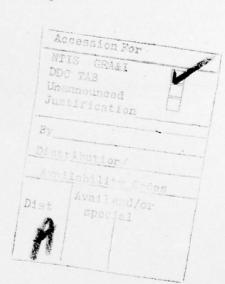| TO DDC-DAA-1 ATTN: Mr. Schrecengost	FROM ARI Rsch Pub Group ~~For the Schreugost +~~ Sre Sc.	DATE 8 Nov 79 Ms Price/48913	CMT 1

1. The reports listed on Inclosure 1 are approved for public release with unlimited distribution (50 numbered ARI Research Memorandums, 74-1 thru 76-30).

2. These are among the previously unrecorded ARI reports which you identified to us 22 June 1979 as not in your retrieval system. The accompanying box contains at least one copy of each report for your retention and reproduction.
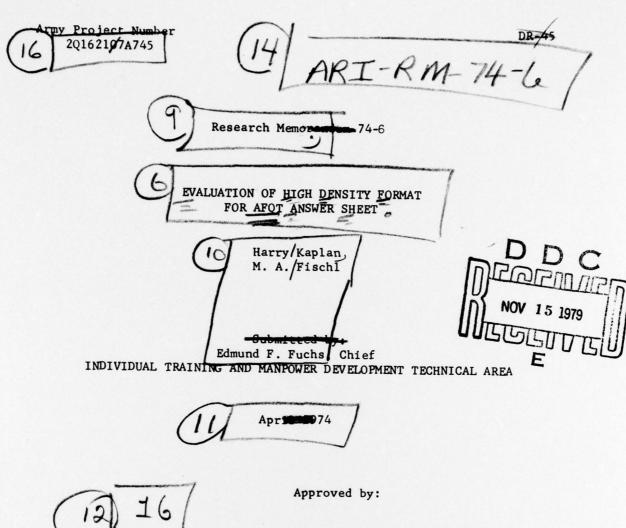
Helen S. Price

1 incl
List of reports, 1974-76

HELEN S. PRICE
Research Publications Group
Army Research Institute

X 4

DA FORM 2496   REPLACES DD FORM 96, WHICH IS OBSOLETE.

DR-45

ARI-RM-74-6

EVALUATION OF HIGH DENSITY FORMAT FOR AFQT ANSWER SHEET

Harry Kaplan
M. A. Fischl

Submitted by:
Edmund F. Fuchs, Chief
INDIVIDUAL TRAINING AND MANPOWER DEVELOPMENT TECHNICAL AREA

April 1974

DDC
RECEIVED
NOV 15 1979
E

Approved by:

E. Ralph Dusek, Director
Individual Training and Performance
Research Laboratory

J. E. Uhlaner, Technical Director
U. S. Army Research Institute for
the Behavioral and Social Sciences

408 010

EVALUATION OF HIGH DENSITY FORMAT FOR AFQT ANSWER SHEET

The Armed Forces Qualification Test (AFQT) was developed jointly by research personnel of the Army, Navy, Air Force, and Marine Corps with the Department of the Army as the executive agent. All the services used the AFQT operationally to determine mental qualifications of male enlistment applicants. The AFQT was also used for screening selective service registrants to determine mental qualification for induction.

The Army Qualification Battery (AQB) is a set of supplementary measures which permitted identification of specific abilities of men marginally acceptable on AFQT or of men who desired to enlist for specific options. Part A of the AQB provides four subtest scores which are obtained by separate scoring of the four content areas of the AFQT.

The DIGITEK Optical Scanner is used to score tests, including AFQT and AQB, at approximately 45 larger Armed Forces Examining and Entrance Stations (AFEES). The DIGITEK is capable of producing several subtest scores and a total score on one pass of the answer sheet through the machine, given an appropriate sequence of items. The sequence of items in AFQT-7C and its alternate form AFQT-8C, which were introduced before the DIGITEK Scanners were installed at the AFEES, is not appropriate for obtaining multiple scores on a single pass. With AFQT-7C and AFQT-8C, five passes are required to obtain the AFQT total score and the four AQB subtest scores.

In order to take advantage of the time-saving multiple scoring feature of the DIGITEK, the items of the AFQT were rearranged to form an experimental test (AFQT-8DX) in which the sequence of items permitted total AFQT and the four Part A AQB subtest scores to be obtained on a single pass. An experimental answer sheet[1] with a revised format was developed for use with the experimental AFQT. It was conceivable that the changes in test and answer sheet format could affect performance on the AFQT and AQB. If, in fact, test performance were to be seriously affected by the changed formats, it would become necessary to restandardize the revised AFQT prior to implementation.

AFQT standardization is a complex process involving a tie-back to a mobilization population. In view of the effort that would be involved in standardization, it was decided that the most appropriate research tactic would be to determine whether standardization was necessary, rather than to standardize automatically.

The ~~general~~ objective of this research was to determine whether format changes in the AFQT test booklet and answer sheet could be introduced operationally without changing existing norms.

1/ PT 4736, Answer Sheet, Armed Forces Qualification Test, AFQT 7DX and 8DX.

Performance on the operational AFQT was compared with performance on an experimental version of the alternate form. Three types of comparability were studied:

(1.) The comparability of test administration difficulty in terms of time, effort expended, examinee understanding, and examinee execution of instructions.

(2.) The comparability of machine scoring efficiency in terms of proportion of answer sheets of each type rejected by the test scoring machine.

(3.) The comparability of scores in terms of means, standard deviations, and correlation coefficients. $2 > p.9$

METHOD

Sampling Procedure

In August 1969, four AFEES were visited by ARI research scientists to initiate experimental test administration. The AFEES were selected to represent a divergent sampling in terms of geographic area and size. This type of selection was made in order to obtain a broadly representative sample, rather than one which would reflect the characteristics of a particular region or community of a particular size within a region.

The following guidelines were used in selecting the sample. Sample size was to be approximately 250 examinees at each installation. The sample would include both Selective Service registrants and applicants for enlistment, but was to be selected at each installation so as to provide as nearly as possible a cross-section of mental ability based on operational AFQT scores. The suggested distribution by mental category was:

Mental Category	Percentage Range
I	5% - 10%
II	25% - 30%
III	30% - 40%
IV	20% - 25%
V	10% - 15%

Samples

The examinees were classified into three samples. Sample A consisted of examinees who took the operational form (AFQT-7C) first and experimental

form (AFQT 8DX) second (7-8 order of administration). Sample B consisted of examinees who took the experimental form first and the operational form second (8-7 order of administration). Sample C consisted of Samples A and B combined.

The number of cases in each sample broken down by testing locations is shown in Table 1.

Table 1

NUMBER OF AFQT EXAMINEES

Testing Location	Sample A 7-8 Order	Sample B 8-7 Order	Sample C A & B
Chicago, IL	156	90	246
Jacksonville, FL	130	125	255
Louisville, KY	201	106	307
New York, NY	119	145	264
	606	466	1,072

Test Administration Procedure

Each examinee in the sample was administered a form of the operational AFQT, using the operational answer sheet, and the alternate form in an experimental format using an experimental answer sheet. A counterbalanced order of administration was attempted at each installation, i.e., one-half of the examinees were to be given the operational form first, followed immediately by administration of the experimental form, and the other half of the examinees were to be given the experimental form first followed immediately by administration of the operational form.

Instruments

Form 7C is a paper-and-pencil multiple choice test consisting of five practice items and 100 test items. The test is in spiral omnibus form with the easiest items at the beginning and the most difficult at the end. There are four content areas occurring in the following sequence of 25 items each: Verbal, Arithmetic Reasoning, Tool Functions, and Spatial Relations. The test is arranged so that groups of four items of each area follow in succession through item 96. The last four items consist of one item from each area. The verbal and arithmetic reasoning items are word items with no illustrations, while the tool function and spatial relations items are picture items with no words. Each picture item consists of five pictures going across the page. A column of four verbal items is followed by a column of four arithmetic reasoning items on the same page. The next page of the sequence contains four tool functions items, and the third page of the sequence consists of four spatial relations items. This sequence repeats itself through item 96. The last page of the test contains one item of each type.

3

The size of the booklet is approximately 10½" x 8". The examinee reads from left to right across the 10½ inch dimension and from top to bottom along the 8 inch dimension. This orientation is less convenient than the conventional one, which is bound on the long dimension and read across the short. The major offsetting advantage is that, although less convenient to handle, this orientation permits larger illustrations than would be possible with the same size conventionally oriented booklet.

The experimental AFQT Form 8DX is an equivalent form to Form 7C, but arranged so that groups of five items of each area follow in succession. The arrangement of items on the pages is the same three-page sequence as in 7C: A column of five verbal items followed by a column of five arithmetic reasoning items on the same page; the next page of the sequence contains five tool functions items, and the third page consists of five spatial relations items. The sequence repeats itself throughout the test.

The size of the booklet is approximately 8" x 10½". In contrast to 7C the examinee reads from left to right across the 8 inch dimension and from top to bottom along the 10½ inch dimension. The conventional orientation of the booklet makes it easier to handle than the 7C booklet. However, in order to fit all five pictures in the picture items across the narrow page dimension, the size of each illustration was reduced by 20%.

The operational answer sheet used with AFQT 7C is DA Form 6010-2, 1 April 1964, Answer Sheet, Armed Forces Qualification Test 7 and 8. This answer sheet has two characteristics which are pertinent to this research: Low density spacing and large letter-block response spaces. In contrast, the experimental answer sheet used with AFQT-8DX (PT 4736, 1 June 1969, Answer Sheet, Armed Forces Qualification Test, 7DX and 8DX) has high density spacing, and small rectangular response spaces. These differences are illustrated in Figure 1.

Variables

The experimental and operational answer sheets were scored in ARI to obtain the following data:

1. AFQT, Form 8DX, Total Score.
2. AQB, VE Score from AFQT-8DX.
3. AQB, AR Score from AFQT-8DX.
4. AQB, SM Score from AFQT-8DX.
5. AQB, PA Score from AFQT-8DX.
6. AFQT Form 7C, Total Score.
7. AQB, VE Score from AFQT-7C.
8. AQB, AR Score from AFQT-7C.
9. AQB, SM Score from AFQT-7C.
10. AQB, PA Score from AFQT-7C.

The test symbols used to designate the AQB subtests scores refer to scores on counterpart tests of the Army Classification Battery (ACB). The

OPERATIONAL ANSWER SHEET FORMAT FOR AFQT-7C

2 A B C D   32 A B C D   62 A B C D   92 A B C D
3 A B C D   33 A B C D   63 A B C D   93 A B C D
4 A B C D   34 A B C D   64 A B C D   94 A B C D
5 A B C D   35 A B C D   65 A B C D   95 A B C D
6 A B C D   36 A B C D   66 A B C D   96 A B C D
7 A B C D   37 A B C D   67 A B C D   97 A B C D
8 A B C D   38 A B C D   68 A B C D   98 A B C D
9 A B C D   39 A B C D   69 A B C D   99 A B C D
10 A B C D   40 A B C D   70 A B C D   100 A B C D

EXPERIMENTAL ANSWER SHEET FORMAT FOR AFQT-8DX

FIGURE 1. COMPARISON OF OPERATIONAL AND EXPERIMENTAL
AFQT ANSWER SHEET FORMATS

verbal subtest is designated VE, the arithmetic reasoning subtest AR, the tool functions subtest SM (for Shop Mechanics), and the spatial relations subtest PA (for Pattern Analysis).

The scores referred to may be raw scores, percentile scores or Army Standard Scores.

Statistical Operations

To determine comparability of machine scoring efficiency, the experimental and operational answer sheets were scored on the DIGITEK Optical Scanner and raw scores were obtained for all variables. Separately for Samples A and B, the different orders of administration, the number of answer sheets selected out by the test scoring machine and the reason for each rejection were tabulated.

Score comparability was determined in two stages. In the first stage, an analysis of variance was performed on the total number of cases to test for the presence of differences between Samples, A vs. B; between test periods, first test vs. second; and between Forms, operational vs. experimental. Raw scores were analyzed separately for AFQT total score and for each of the four subtest scores, utilizing a Latin Square design with repeated measures on the same subjects over the Forms and periods variables. Computations were performed on the unequal size subject groups utilizing the unweighted means method.

In the second stage, the total sample (Sample C) was stratified on the basis of total AFQT 7C percentile scores to be representative of the AFQT mobilization population. Stratification was accomplished by using all cases and weighting the frequency in each decile by a multiplier such that the effective frequencies in all deciles were equal.

In the stratified sample, the following computations were made:

1. Raw score means and standard deviations for all variables.

2. An intercorrelation matrix for all variables.

3. Cumulative percentiles for AFQT-7C and AFQT-8DX.

RESULTS

Observations of the administration of the experimental vs. the operational AFQT revealed no administrative difficulties peculiar to the experimental form.

Thirteen experimental and fifteen operational answer sheets were selected out by the DIGITEK optical scanner. This result demonstrated that efficiency of machine scoring was not impaired by the new format.

The analysis of variance tables used in determining sample, test period, and form differences are shown in the Appendix (Tables A-1, A-2, A-3, A-4, and A-5). Means of the various factors and significance of the differences between these means are summarized in Table 2. These results show a statistically significant difference between the two forms on AFQT total score. Results of analysis of the subtest scores show these form differences to be contributed to by statistically significant differences in the SM and PA subtests, with no form differences in VE and AR subtests. The samples also differed in SM score, test periods in PA score, with no other statistically significant differences being detected.

The significant differences occurred in the picture items and not in the word items. This occurrence is thus a function of changes made to the test booklet, not the revised answer sheet format. Since the picture items in the experimental booklet are considerably smaller than those in the operational booklets, it was concluded that reducing the size of the pictures increased the difficulty of the picture items. The obvious solution to this problem is to enlarge the size of the pictures in the experimental forms.

Pooling over the individual sample and test period differences (statistically significant but practically small), the total sample was stratified, and these stratified sample means, standard deviations and intercorrelation coefficients are shown in Table 3.

The correlation of .95 between the AFQT scores in the two forms is a strong indication that the two forms remain alternate despite change in format of booklet and answer sheet (the original standardization r was .93). The correlation of each subtest with its alternate form subtest (.93, .92, .86, and .86 for VE, AR, SM, and PA, respectively) are equally acceptable. Thus, change in size of pictures is the only change considered necessary.

Table 2

SUMMARY OF UNWEIGHTED MEAN SCORE DIFFERENCES

| | Samples | | Test Periods | | | Forms | | |
	A	B	1st Test		2d Test	7C		8DX
AFQT	58.42 NS	56.60	57.06	.01	57.96	58.23	.01	56.79
VE	16.41 NS	15.98	16.20	NS	16.19	16.18	NS	16.21
AR	15.14 NS	14.94	15.05	NS	15.03	14.94	NS	15.14
SM	13.66 .01	12.61	13.03	NS	13.24	13.60	.01	12.67
PA	13.44 NS	13.31	12.98	.01	13.77	13.71	.01	13.04

Table 3

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATION COEFFICIENTS OF AFQT 7C
AND AFQT 8DX TOTAL AND SUBTEST SCORES FOR STRATIFIED SAMPLE

N = 1072

TEST	Raw Score MEAN	SD	Inter - r's									
			VE 7C	AR 7C	SM 7C	PA 7C	AFQT 7C	VE 8DX	AR 8DX	SM 8DX	PA 8DX	AFQT 8DX
VE - 7C	15.43	8.50	VE 7C									
AR - 7C	14.33	8.33	.83	AR 7C								
SM - 7C	12.97	5.96	.51	.53	SM 7C							
PA - 7C	13.03	8.16	.62	.69	.60	PA 7C						
AFQT-7C	55.60	27.62	.88	.90	.75	.85	AFQT 7C					
VE 8DX	15.49	8.73	.93	.83	.50	.62	.86	VE 8DX				
AR 8DX	14.50	8.45	.81	.92	.51	.67	.86	.82	AR 8DX			
SM 8DX	12.11	7.07	.46	.47	.86	.54	.66	.46	.47	SM 8DX		
PA 8DX	12.51	8.08	.64	.71	.58	.86	.82	.66	.72	.55	PA 8DX	
AFQT 8DX	54.37	27.67	.85	.88	.71	.80	.95	.88	.90	.71	.87	AFQT 8DX

CONCLUSIONS

1. Use of the high density answer sheet does not make the AFQT more difficult than use of the low density answer sheet. The high density AFQT answer sheet may be substituted for the low density AFQT answer sheet in operational programs without restandardization.

2. Reducing the size of the illustrations appears to increase the difficulty of the tool functions and spatial relations subtests.

Before the test format in the AFQT 8DX booklet is used operationally, the illustrations should be enlarged. A 10% reduction from the size contained in the AFQT 7C booklet is the smallest size considered adequate.

9

APPENDIX

Table A-1

ANALYSIS OF VARIANCE: TOTAL SCORE DERIVED FROM AFQT 7C AND AFQT 8DX

Source	Sums of Squares	DF	Mean Squares	F	P
<u>Between Subjects</u>		<u>1071</u>			
Samples	1,739.0	1	1,739.0	1.51	NS
Subjects within samples	1,229,613.0	1070	1,149.2		
<u>Within Subjects</u>		<u>1072</u>			
Test Periods	410.0	1	410.0	10.85	.01
Forms	1,085.0	1	1,085.0	28.70	.01
Error	40,449.0	1070	37.8		

Table A-2

ANALYSIS OF VARIANCE: VERBAL (VE) SUBTEST SCORE DERIVED FROM AFQT 7C AND AFQT 8DX

Source	Sums of Squares	DF	Mean Squares	F	P
Between Subjects		1071			
Samples	97.5	1	97.5	0.81	NS
Subjects within samples	128,822.0	1070	120.4		
Within Subjects		1072			
Test Periods	0.0	1	0.0	0.01	NS
Forms	0.6	1	0.6	0.09	NS
Error	6,280.2	1070	5.9		

Table A-3

ANALYSIS OF VARIANCE: ARITHMETIC REASONING (AR) SUBTEST SCORE DERIVED FROM AFQT 7C AND AFQT 8DX

Source	Sums of Squares	DF	Mean Squares	F	P
Between Subjects		1071			
Samples	21.0	1	21.0	0.18	NS
Subjects within samples	123,861.0	1070	115.8		
Within Subjects		1072			
Test Periods	0.2	1	0.2	0.03	NS
Forms	21.0	1	21.0	3.31	NS
Error	6,794.7	1070	6.4		

12

Table A-4

ANALYSIS OF VARIANCE: SHOP MECHANICS (SM) SUBTEST SCORE DERIVED FROM AFQT 7C AND AFQT 8DX

Source	Sums of Squares	DF	Mean Squares	F	P
Between Subjects		1071			
Samples	581.0	1	581.0	8.03	.01
Subjects within samples	77,436.0	1070	72.4		
Within Subjects		1072			
Test Periods	23.7	1	23.7	3.17	NS
Forms	456.3	1	456.3	60.00	.01
Error	8,126.0	1070	7.6		

Table A-5

ANALYSIS OF VARIANCE: PATTERN ANALYSIS (PA) SUBTEST SCORE DERIVED FROM AFQT 7C AND AFQT 8DX

Source	Sums of Squares	DF	Mean Squares	F	P
Between Subjects		1071			
Samples	9.3	1	9.3	0.07	NS
Subjects within samples	113,803.0	1070	106.4		
Within Subjects		1072			
Test Periods	334.5	1	334.5	34.61	.01
Forms	243.3	1	243.3	25.18	.01
Error	10,339.5	1070	9.7		

13